COVFILTER: A LOW-COST PORTABLE DEVICE FOR THE PREDICTION OF COVID-19 FOR RESOURCE-CONSTRAINED RURAL COMMUNITIES

Sajedul Talukder[1] and Faruk Hossen[2]

[1]School of Computing, Southern Illinois University, USA
sajedul.talukder@siu.edu
[2]Department of Computer Science, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh
faruk.08.cse@gmail.com

ABSTRACT

Early identification of COVID-19 is critical for preventing death and significant illness. People living in remote parts of resource-constrained countries find it more difficult to get tested due to a lack of adequate testing. As a result, having a primary filtering tool that can assist us in simplifying bulk COVID testing to prevent community spread is vital. In this paper, we introduce CovFilter, a low-cost portable device for COVID-19 prediction for resource-constrained rural communities, with the goal of encouraging people to be tested for COVID-19 in a more informed manner. CovFilter Hardware Module collects health parameters from three sensors while the CovFilter Prediction Module predicts COVID-19 status using the health data. We train supervised learning algorithms and an artificial neural network to predict COVID-19 from vital sign readings where MultilayerPerceptron outperformed ANN, NaiveBayes, Logistic, SGD, DecisionStump, and SVM with an F1 of 93.22%. We further show that a weighted majority voting ensemble classifier can outperform all single classifiers achieving an F1 of over 94%.

KEYWORDS

COVID-19 prediction, CovFilter, Arduino, Machine learning, Rural community.

DOI: 10.5121/ijaia.2022.13201

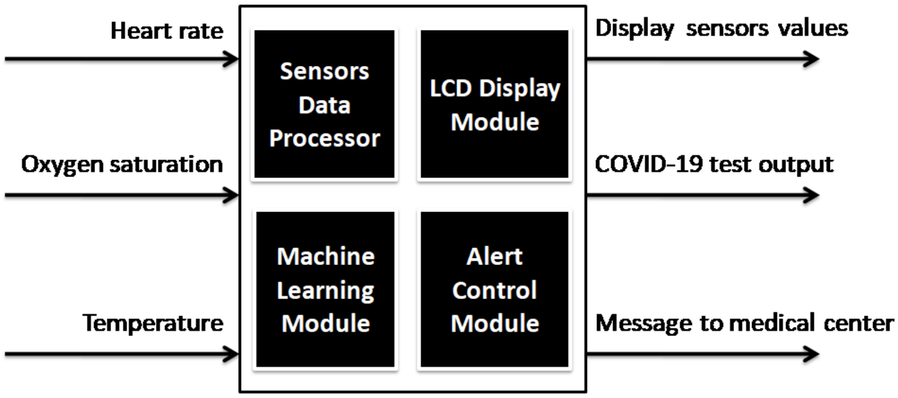

Figure 1. System diagram.

1. INTRODUCTION

COVID-19, a disease spread by the SARS-CoV-2 virus, has ravaged the whole world, even resource-strapped third-world countries [1]. Infected people might experience anything from mild to moderate respiratory illness to being critically ill and requiring hospitalization. Elderly people and those with sophisticated medical constraints like diabetes, cardiovascular disease, chronic respiratory disease, or cancer are probable to evolve into serious medical emergencies. People of all ages may become sick with COVID-19 and feel seriously unwell or die. The Coronavirus can spread from a victim's mouth or nose in tiny infectious particles when they speak, cough, sneeze, breathe. However, detection of COVID-19 is hard as the patients express nonspecific symptoms ranging from cough (68%), fatigue (38%), sputum production (34%), and breathlessness (19%), and an identical method cannot be used for an accurate diagnosis [2]. A lot of these symptoms might be related to other respiratory infections. Only prompt testing to detect COVID-19 and segregation of patients from the general population to minimize viral spread is the viable option to reduce transmission and put an end to COVID-19. COVID-19 must be identified as soon as possible to avoid death and serious illness.

While the epidemic has affected every community, it has also highlighted significant socioeconomic and healthcare disparities, particularly among rural residents. Even in the United States, eighty percent of the population in remote locations is classified as medically underserved [3].

Command	Description
AT	Ensure that system is working fine.
AT+CSCA?	Takes the source number as input before send message command.
AT + CMGF = 1	Chooses the message format.
AT + CMGS = “88001743. . . .78”	Takes destination number and sends the message to destination.
1A	It is called terminator. After sending the message, this terminator has to indicate that sending is successful.

Table 1. GSM commands to create and send text messages

Furthermore, remote places have greater rates of disability and inhabitants without health-care insurance or broadband internet connection, restricting telemedicine availability. According to a Surgo Foundation analysis, more than two-thirds (64%) of all rural counties in the United States lack a COVID-19 testing center, leaving 20.7 million people without access to testing [4]. If this is the state of the world’s wealthiest nation, we can readily envisage the state of resource-constrained impoverished countries. Effective public health response to COVID-19 is hampered by a lack of rural health infrastructure and restricted access to health services. The closure of hospitals owing to greater numbers of uninsured and lower-income inhabitants, outmoded payment mechanisms, aging populations, and older infrastructure are all contributing factors to the loss of health care in rural America. In many rural locations, there is a dearth of health professionals for both basic and specialist treatment. Moreover, people living in rural areas are unaware of COVID-19’s heinous nature and refuse to test it. This is because they must travel to a test center often located in a city and adhere to certain formalities in order to test COVID-19, which they consider to be a waste of time. As a result, identifying a patient can take a long time, and if treatment is not started promptly, the patient may die. In the presence of highly transmissible variants like Delta and Omicron, it is critical that we have a primary filtering mechanism that can help us to simplify mass COVID testing to prevent community spread.

In this paper, we present a cost-effective approach to assist rural areas in diagnosing COVID-19 early with the help of low-cost portable hardware. Our system has two key components: the hardware module consisting of sensors module, LCD module, and GSM module that collects health parameters from wearable sensors while the machine learning module predicts COVID-19 using the data readings collected by the hardware module (see Figure 1).

2. OBJECTIVE

Rural communities work hard to make a livelihood, but they don’t care in the slightest about their health or medications. As a result, bringing people under fast COVID-19 testing is a source of concern and a major burden for the government. Rural communities, moreover, have significant challenges in accessing health care due to limited or non-existent testing resources. To solve the problem, we need a system that is vital for impoverished, underprivileged, and rural areas to alleviate the situation. Our goal is to build a system that is cost-effective, time-consuming, and easy to use must be developed, allowing for quick COVID-19 prediction. COVID-19 would not be precisely identified, but the approach would estimate the chance of infection with COVID-19. To be certain, someone who is most likely infected with Coronavirus (SARS-CoV-2) should self-quarantine, continue preparatory therapy, and undergo COVID-19 testing at a testing facility. In this regard, the designed method should serve as a key filtering device for the risk of COVID-19 infection in rural areas. In particular, aim to achieve the following key objectives when designing our system:

- **Low-cost:** Our system should be as cost-effective as possible since the goal is to design a system that will serve the impoverished rural population. To be more exact, using the system as many times as needed should cost the very minimum or nothing at all.
- **Simple:** Since the device will be used by individuals with limited technological understanding, one significant goal is to create a system that does not require any technical knowledge to use. Another consideration is to have a straightforward operational interface that is easy to understand and requires little previous training.
- **Energy-efficient:** Another criterion is that the gadget requires little power to function. We can't design a system that would possibly be costly due to its high energy consumption.
- **Battery or solar-powered:** Because many rural locations lack access to electricity, the gadget must be able to run on battery or solar power.
- **Portable:** Due to lack of access to a good transportation system, it is often hard to move from one place to another and travel long distances. Furthermore, rural regions, which have higher percentages of handicapped and aging populations, require a system that is portable and easy to transport to the households of vulnerable patients.

3. CovFilter System

Rapid and thorough coronavirus (SARS-CoV-2) detection is important for containing the COVID-19 pandemic. A good health policy implementation requires early identification of ill people and substantial contact tracing. Our solution would largely be installed in rural pharmacies, allowing

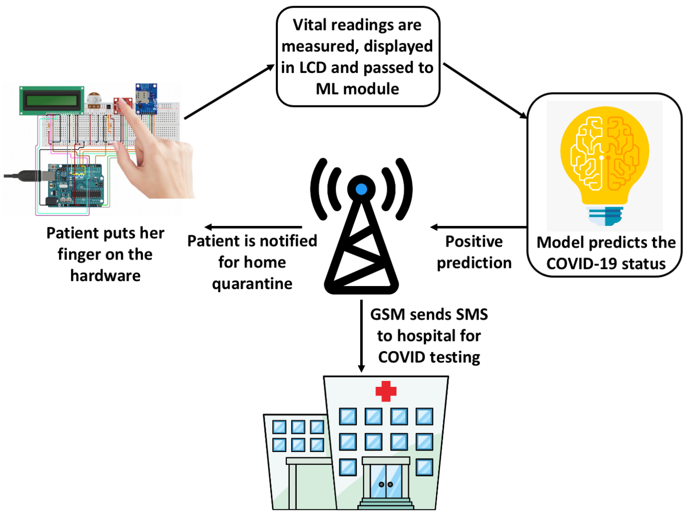

Figure 2. CovFilter System.

thousands of people to benefit from it. It would be simple and convenient for rural areas to obtain a COVID-19 prediction, and the cost of the test would be quite low. Our system is composed of two key components: a microcontroller-based hardware module and a machine learning module. A Hardware system fetches different health parameters of an individual by using a micro-controller and sensors attached within that system.

The temperature sensor and the pulse oximeter senses the temperature, pulse rate, and oxygen saturation of an individual. The 16x02 LCD module shows the system's measured temperature, pulse rate, and oxygen saturation. The system is built using hardware components that output temperature, heart rate, and oxygen saturation to input a machine learning module that predicts the COVID-19 of an individual. If the COVID-19 prediction is affirmative, the alert control module sends a message to the COVID-19 testing center, requesting that a real test be scheduled. It also sends a communication to the local health department, alerting them to a possible new COVID case and putting them on home quarantine. A basic overview of the system is shown in Figure 2.

3.1. Hardware Components

To build our system, we have used the following hardware components: Micro-controller (Arduino UNO), Temperature sensor (DS18B20 1-Wire), Pulse oximeter (MAX30100), LCD (16x2 LCD), GSM Module (SIM800L), Breadboard, Resistor, Wires, etc. Figure 3 shows the circuit implementation of our hardware module.

3.2. Temperature Sensor (DS18B20 1-Wire)

The DS18B20 temperature sensor is very precise and works without external components. It is possible to measure temperatures from -55°C to +125°C with ±0.5°C accuracy with this sensor. The resolution of the temperature sensor is configurable. Users can configure resolution up to 9, 10, 11, or 12 bits but the default resolution at power-up is 12-bit. DS18B20 1-Wire is a digital temperature sensor. It has three terminals such as red, black and yellow. The red terminal indicates VCC, the black terminal indicates GND, and the yellow terminal indicates digital input. It is interfaced with Arduino as follows- VCC is connected to 5V PIN of Arduino, GND is connected to GND PIN of Arduino and digital input PIN is connected to PIN 2 of Arduino.

3.3. GSM Module

GSM stands for Global System for Mobile which is a second-generation (2G) cellular standard developed to cater to voice services and data delivery. The main purpose of GSM is calling, sending SMS, and receiving SMS. In the system, GSM is used only for sending SMS. Table 1 shows the commands that can be used to control the output of the GSM module for a text message. Major components of this module are power, antenna, SIM slot, RX (Receiver) PIN, TX (Transmitter) PIN, etc.

Interfacing of GSM with Arduino is as follows- VCC of GSM to 5V of Arduino, GND of GSM to GND of Arduino, TXD PIN to PIN10 of Arduino, RXD PIN of GSM to PIN 11 of Arduino. A SIM card is inserted into the SIM slot.

3.4. Pulse Oximeter

The pulse oximeter is used as a key health indicator that measures oxygen saturation (percentage of SpO2 concentration in blood) and pulse rate. Light sources and photo-sensor are incorporated to determine SpO2 concentration. In modern days, hospital equipment includes a SpO2 module to measure the availability of oxygen in the blood of a patient. In this study, we have used MAX30100 based pulse oximeter. Oxygen saturation is measured without pain where oxygen saturation is the level of oxygen sent out from the heart to all parts of the body, such as arms and legs. It is used to test whether enough oxygen is available in the blood and to check the condition of health of a person.

The pulse oximeter module includes light that is used in measuring the pulse rate. To compute pulse rate one has to place his finger on the pulse oximeter. Based on the volume of blood in the capillary blood vessels, the light reflected changes. The volume inside the capillary blood vessels will be high during a heartbeat. This impact on the reflection of light and the reflection during heartbeat will be less compared to the reflection during no heartbeat. This variation in light reflection can be gathered as a pulse rate from the output channel of the oximeter. This pulse can be shaped to compute heartbeat and then programmed appropriately to read as heartbeat count to transform into pulse rate.

The pulse oximeter is connected to Arduino as the following: 5V of the pulse oximeter to 5V of Arduino, GND of the pulse oximeter to GND of Arduino, SDA PIN of the pulse oximeter to A4 PIN of Arduino, and SCL PIN of the pulse oximeter to A5 PIN of Arduino.

3.5. 16x2 LCD

Liquid Crystal Display (LCD) is an electronic display module that utilizes liquid crystals to create a visible image. The 16×2 LCD module is used in circuits to display text and computed values. This module displays 16 characters per line in two lines. Each character is displayed in a 5×7 pixel matrix in this type of LCD.

LCD has two types of PINs such as control PINs and data PINs. Interfacing of control PINs of 16x2 LCD with Arduino is VO to changeable POT that can supply from 0 to 5V, VSS to GND of Arduino, VDD to 5V of Arduino, RS to PIN 8 of Arduino, E to PIN 7 of Arduino, RW to GND of Arduino, K to GND of Arduino. Data pins of LCD interfaced with Arduino as D4 to PIN 3 of Arduino, D5 to PIN 4 of Arduino, D6 to PIN 5 of Arduino, D7 to PIN 6 of Arduino.

4. DATASET

In the experiment, we use the COVID-19 Temperature, Oxygen & Pulse Rate Readings Dataset from Kaggle [5]. The dataset consists of momentary vital readings with a label indicating if the patient was Positive or Negative for COVID-19. The data is mostly synthetic and artificially generated data based on the initial real-world readings. The COVID-19 dataset includes 10000 instances, 3 features, and 2 target classes for COVID-19 status; either positive or negative. The dataset has no missing values. In this dataset, the features describe the momentary vital sign readings. As described in the dataset, three real-valued features represent the temperature, oxygen

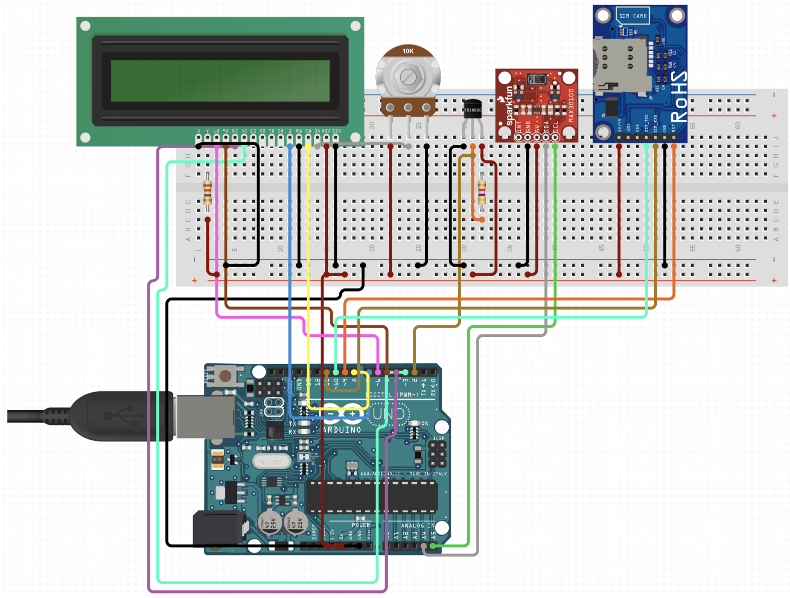

Figure 3. System Circuit: temperature sensor, pulse oximeter, and GSM module are interfaced with Arduino. LCD is used in the circuit to display text and computed values.

saturation & pulse rate readings. The dataset is mostly balanced with 4990 positive instances and 5010 negative instances.

5. CovFilter Prediction Module (CPM)

We have used 10-fold cross-validation to evaluate the ability of the CovFilter prediction module (CPM) to predict the COVID-19 status of the patients from the vital signs readings data. We have divided this dataset into 10 folds of 1000 tuples each, selected randomly, and used 10-fold cross-validation to evaluate several supervised learning algorithms. That is, in each of the 10 experiments, we used 9 folds to train and one to test. We used the features of each tuple in the 9 folds to separately train supervised learning algorithms. Then, for each tuple in the remaining fold, CPM uses the trained algorithms to predict the user's COVID-19 prediction. We report weighted averages of the prediction accuracy over the 10 experiments.

We have used Weka 3.8.5[1] to evaluate several supervised learning algorithms, including Decision Stump (DS), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Multinomial Logistic Regression (Logistic), Multilayer Perceptron (MP), and Naive Bayes (NB). Finally, we trained a weighted majority voting ensemble classifier using Weka.classifiers.meta.vote where we selected Majority Voting as a combination rule. In our ensemble, we used a combination of Mul-

[1]https://www.cs.waikato.ac.nz/ml/weka/.

tilayer Perceptron, Support Vector Machine, Multinomial Logistic Regression, and Decision Tree. Each of these classifiers predicts a nominal class label for a test sample. The label which is predicted the most then gets selected as the output of the Majority Voting (MV) classifier. All nominal classes of the test instance are loaded into votes. Each classifier classifies the instance and the label with the highest probability gets a vote. If multiple labels have the same probability then all these labels receive a vote. Once all classifiers have cast their vote, the label with the most votes is selected as the label for the test instance. If multiple labels have the same amount of votes, then one of these labels will randomly be selected.

We have also used Artificial Neural Network (ANN) which is a deep learning model. The ANN is constructed by adding three hidden layers in the network along with the input layer, and output layer. The three hidden layers have 32, 64, and 1 neuron respectively where the first two layers use relu activation and the final one uses sigmoid activation. To train the network it uses the back-propagation method to minimize the squared reconstruction error. We have compiled the model with Adam optimizer which is an alternative optimization algorithm that provides more efficient neural network weights by running repeated cycles of adaptive moment estimation. To calculate the loss, we have used binary cross-entropy that compares each of the predicted probabilities to actual class output which can be either 0 or 1.

6. RESULT ANALYSIS

To evaluate the robustness of the classifiers, we performed 10-fold cross-validation and results are presented in terms of the average classification accuracy. We have used the following metrics to demonstrate the feasibility and applicability of the proposed CovFilter prediction module (CPM): (1) *Performance*, (2) *False Positive Rate*, and (2) *Mean Absolute Error and Root Mean Squared Error*.

Performance. Our system predicts a binary classified outcome that outputs the class value into either of the two classes: COVID negative and COVID positive. We have measured the performance of our system through Precision, Recall, and F-measure. An increase in those parameters indicates that the system is performing better. The performance parameter values for both negative and positive classes of the system are shown in Figure 4.

False Positive Rate (FPR). False Positive Rate indicates performance degradation of the system. FP means mistakenly identifying a person as COVID positive which is very much disagreeable. An increase in FPR means a decrease in performance. So measurement of FPR is an aspect of

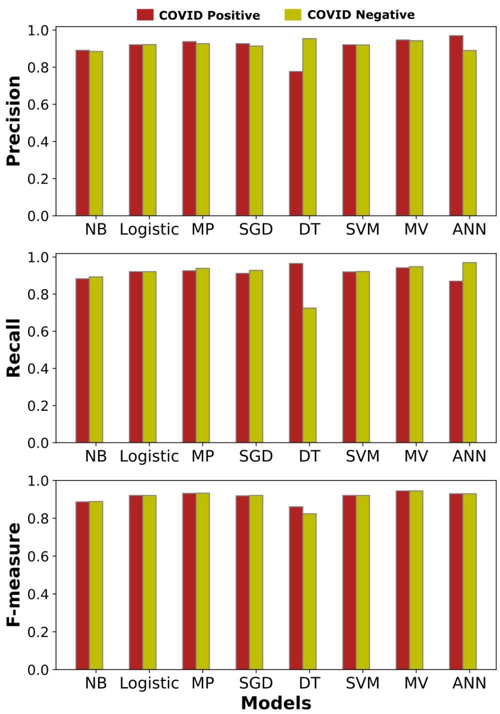

Figure 4. Precision, Recall and F-measure for the positive and negative classes obtained from various classifiers. MajorityVoting outperforms all the other models.

evaluating a system. Figure 5(a) shows FPR for negative and positive result.

Mean Absolute Error and Root Mean Squared Error. Mean absolute error (MAE) is the mean of a measure of errors between a paired error of input and output. Root mean squared error (RMSE) indicates the square root of the mean of the square of all errors. An increase in any of these errors means a decrease in performance. We have computed these errors to evaluate our system. Figure 5(b) shows the mean absolute error and root mean squared error of all the models implemented in the system.

6.1. Evaluation of the Classifiers

We trained and tested our classifier models using data from the COVID-19 Temperature, Oxygen & Pulse Rate Readings Dataset. For the prediction task, our dataset consists of tuples (X_i, Y_i), where X_i is the feature vector and Y_i is the COVID-19 prediction.

Table 2 shows the Precision, Recall, and F1 measure for different machine learning and deep

Algorithm	Kappa	FP Rate	Precision	Recall	F1
NaiveBayes	0.776	11.2%	88.8%	88.8%	88.8%
Logistic	0.8426	7.9%	92.1%	92.1%	92.1%
MultilayerPerceptron	**0.8644**	**6.8%**	**93.2%**	**93.2%**	**93.2%**
SGD	0.8404	8.0%	92.0%	92.0%	92.0%
DecisionStump	0.6895	15.5%	86.6%	84.5%	84.2%
SVM	0.8418	7.9%	92.1%	92.1%	92.1%
MajorityVoting	**0.8904**	**5.5%**	**94.5%**	**94.5%**	**94.5%**
Artificial Neural Network (ANN)	0.8506	7.9%	93%	93%	93%

Table 2. Kappa statistic, False Positive rate, Precision, Recall, and F1-measure for different machine learning classifiers where MultilayerPerceptron achieves the best performance (F1: 93.22%, precision: 93.22% and recall: 93.22%). However, a weighted majority voting ensemble achieved an F1 of 94.5%, with a precision of 94.5% and a recall of 94.5%.

learning classifiers along with their Kappa statistic and False Positive rate. MultilayerPerceptron outperformed NaiveBayes, Logistic, SGD, DecisionStump, and SVM. Specifically, MultilayerPerceptron achieved an F1 of 93.22%, with a precision of 93.22% and a recall of 93.22%. ANN achieved an F1 of 93%, with a precision of 93% and a recall of 93%. However, the weighted majority voting ensemble classifier outperforms all achieving an F1 of 94.5%, with a precision of 94.5% and a recall of 94.5%.

6.2. Feature Rank

The most informative feature in terms of information gain using ranking filter was oxygen saturation reading (0.6254), whereas the least informative feature was Pulse Rate (0.0205). We did not find any significant positive or negative correlations between the features; pulse rate and oxygen (Pearson correlation coefficient of -0.0057), temperature and oxygen (Pearson correlation coefficient of -0.0156), and temperature and pulse rate (Pearson correlation coefficient of 0.0096).

7. Related Work

Machine learning-based solutions are becoming increasingly popular, with applications range of social networks [6–10] to e-governance [11, 12], from digital automation [13, 14] and cellular networks [15] to cybersecurity [16, 17]. There have been several attempts to detect COVID-19 using machine learning and make an early prediction on the symptoms. The works that relate to our approach fall into two major categories: COVID prediction using machine learning and

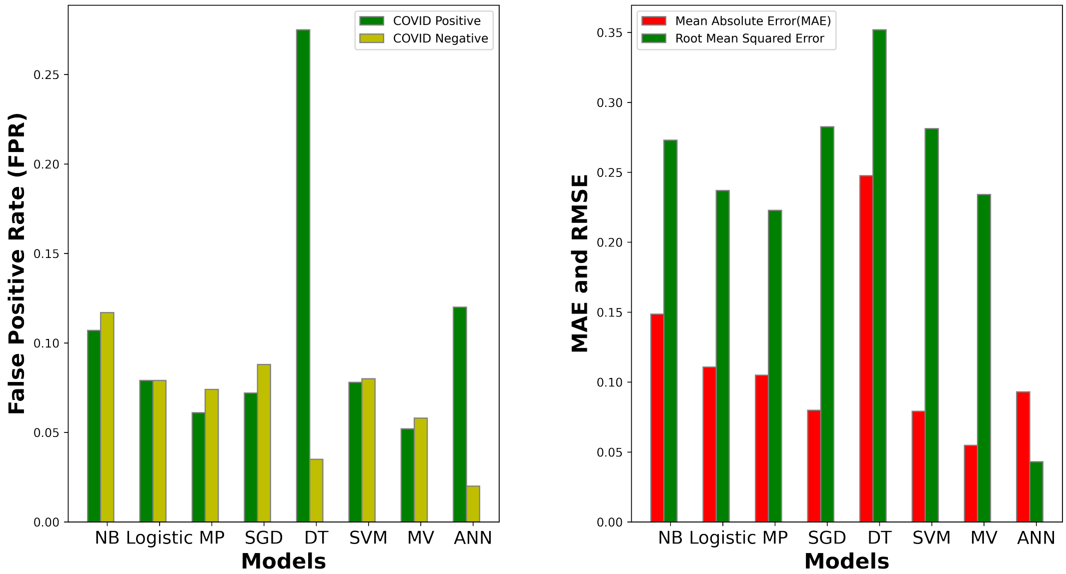

Figure 5. (a) False Positive Rate (FPR) of all models. FPR is an important aspect of evaluating a system. Among all models, Majority voting has the lowest FPR and we the get best performance. (b) Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for all models. ANN has the least MAE and RMSE among all models.

hardware solutions for early COVID detection. We detail the related work below.

7.1. COVID prediction using machine and deep learning

To diagnose COVID-19 early, Vangipuram et al. developed a machine learning-based health monitoring framework for the diagnosis of infected patients [18]. Darapaneni et al. applied the classification method of supervised learning on the laboratory tests datasets to predict COVID-19 [19]. SEIR and Regression-based machine learning models were used to predict the rate of spread of COVID-19 so that government and doctors can prepare the future plan to diminish the spread of the virus [1, 11]. Hallman et al. developed a predictive model that would help clinical personnel in the diagnosis of COVID-19 patients and suggest medical care for patients with either semi-intensive or intensive care [20]. In another work, features were drawn from six images which are acquired from thermal video clips and classified using a conventional machine learning model to predict COVID-19 [21]. Polynomial regression method of machine learning had also been applied that aims at predicting COVID-19 under limited data conditions to prevent the outbreak of the virus [22]. Zervoudakis et al. used the Bayesian machine learning model to train data set related to symptoms and social life and predict COVID-19 infected patients [7]. There has been an attempt to apply machine learning techniques on coughs and breathing sounds to classify healthy and COVID-19 sounds [23]. Feng et al. developed a method for diagnosis of COVID-19 through cough sound extraction, sound feature extraction, cough detection, and cough classification [24]. COVID-19 was predicted on cough audios using the logistic regression method where features are extracted from cough audios [25].

Several convolutional neural networks (CNNs), involving VGG16, ResNet50, DenseNet121, and InceptionResNetV2 extracted features from chest X-ray and computed tomography (CT) images, then different machine learning techniques were applied on extracted features to identify COVID-19 cases [26]. Fedal et al. proposed a parallel Convolutional Neural Networks (CNNs) model for the detection of COVID-19 infected patients using chest X-ray radiographs [27]. Khan et al. presented a 3D-deep learning-based method that automatically screens coronavirus patients using 3D volumetric CT image data where CT is Computed Tomography (CT) image of lung [13]. Large scale test of COVID-19 so costly for people in rural areas, so an intelligent clinical decision support system (SADC) was developed that extracts features from chest x-rays and predicted COVID-19 using deep learning [28]. Rahman et al. developed a suite of deep neural network (DNN) based COVID-19 case detection and recognition framework [29]. Neural networks and transfer learning were incorporated to predict COVID-19 infectees through radio-graphs of chest X-rays [30]. Convolutional Neural Network (CNN) along with the existing famous pre-trained networks were employed for prediction of COVID 19 [30]. Zokaeinikoo et al. suggested an artificial intelligence model for the detection of COVID-19 (AIDCOV) that classified chest radiography images of individuals either COVID-19, other infections, or normal [31]. Kumar et al. evolved an Auto-Regressive Integrated Moving Average and Long Short-Term Memory (LSTM) model for predicting the cumulative number of COVID-19 cases and the cumulative number of deaths [32]. According to Sevi et al. applying the augmentation method to the data set extracted from X-rays, patients were classified through multi-class classification deep learning models [33]. Demystify technique was applied to detect COVID-19 assembling medical images with the help of deep nets [34]. Li et al. presented a method to detect and classify COVID-19 based on the YOLOv5 model [35]. Shadeed et al. created a system to identify and detect COVID-19 disease with the help of X-ray radiation, GoogLeNet, ResNet-101, Inception v3 network, and DAG3Net [36]. According to Goswami et al., local binary pattern-based feature selection along with a convolutional neural network was applied to predict COVID-19 analyzing chest X-ray image [37]. Kharbat et al. incorporated artificial intelligence and data mining techniques for diagnosing COVID-19 from chest X-ray images [38]. Huang et al. combined a belief function-based convolutional neural network with semi-supervised training to detect COVID-19 cases [39].

7.2. Hardware solutions for early COVID detection

Xian et al. proposed an electrochemical detection system for SARS-CoV-2 and cardiac troponin I with the help of field-effect transistor (FET) based biosensor hardware with low cost [40]. A Micro-controller (Arduino) based HW/SW system was developed that was cost-effective compared to OEM solutions and applied in wearables, telemedicine, and IoT healthcare systems in process of COVID-19 diagnosis and treatment [41]. Quer et al. developed a smartphone app that differentiates COVID-19 positive versus negative cases based on smartwatch and activity tracker data, self-reported symptoms, and diagnostic testing results from individuals and sensor data [42]. Stojanović et al. utilized sensors, microcontrollers, gadgets, and peripheries to detect and monitor body temperature, heart rate, respiration rate, and other vital signs, then measured health parameters were correlated to COVID-19 or similar diseases [43]. Natarajan et al. proposed methods that collected symptoms from wearable devices and symptoms reported by individuals then measured prediction of hospitalization applying logistic regression based on symptoms [44]. Mishra et al. developed an online detection algorithm to identify early stages of infection COVID-19 by real-time heart rate from smartwatch [45]. Un et al. utilized wearable biosensors and machine learning-based remote monitoring platforms for managing COVID-19 patients hospitalized in isolation wards where the machine learning method was applied physiology parameters taken from wearable biosensors, symptoms, and other medical data [46]. As stated by Zhu et al., smartwatches were used to extract symptoms to identify and predict COVID-19 infection [47]. Bukkitgar et al. suggested an electrochemical method in conjunction with nanotechnology for the detection of SARS-CoV-2 [48]. Surya and Yarlagadda developed an artificial intelligence-based smart device that is economical to detect COVID-19 and alerts people about maintaining social distance [49].

8. CONCLUSIONS

We introduce CovFilter, a low-cost portable device for COVID-19 prediction in resource-constrained rural populations, in this study. The hardware subsystem gathers health metrics from sensors, while the COVID classifier subsystem predicts COVID-19 based on the health values acquired by the hardware subsystem. We trained supervised learning algorithms to predict COVID-19 from vital sign data, and MultilayerPerceptron beat NaiveBayes, Logistic, SGD, DecisionStump, and SVM with an F1 of 93.22%. We also demonstrate that a weighted majority voting ensemble classifier outperforms all single classifiers, with an F1 of 94.5 percent. Our system has the potential to be used as a wearable household device to predict COVID-19 from the comfort of one's own home.

REFERENCES

[1] A. S. Shonchoy, K. S. Ishtiaq, S. Talukder, N. U. Ahmed, and R. Chowdhury, "A novel index-based decision support toolkit for safe reopening following a generalized lockdown in low and middle-income countries," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[2] W.-J. Guan, Z.-y. Ni, Y. Hu, W. Liang, C.-Q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. Hui, B. Du, L.-j. Li, G. Zeng, K.-Y. Yuen, R.-c. Chen, C.-l. Tang, T. Wang, P.-y. Chen, J. Xiang, and N.-s. Zhong, "Clinical characteristics of coronavirus disease 2019 in china," *New England Journal of Medicine*, vol. 382, 02 2020.

[3] "Covid-19 policy brief: Disparities among rural communities in the united states." shorturl.at/hBMN8, December 2020.

[4] B. Hardy, "New surgo analysis identifies highly vulnerable rural communities as covid-19 testing deserts." shorturl.at/gGQV3, June 2020.

[5] R. Mascarenhas, "Covid-19 temperature, oxygen pulse rate readings," 2021.

[6] S. Talukder and B. Carbunar, "A study of friend abuse perception in facebook," *Transactions on Social Computing*, vol. 1, no. 1, 2020.

[7] S. Zervoudakis, E. Marakakis, H. Kondylakis, and S. Goumas, "Prediction of covid-19 infection based on symptoms and social life using machine learning techniques," in *The 14th PErvasive Technologies Related to Assistive Environments Conference*, PETRA 2021, (New York, NY, USA), p. 277–283, Association for Computing Machinery, 2021.

[8] S. Talukder and B. Carbunar, "AbuSniff: Automatic Detection and Defenses Against Abusive Facebook Friends," in *Twelfth International AAAI Conference on Web and Social Media*, 2018.

[9] S. K. Talukder, "Detection and prevention of abuse in online social networks," *FIU Electronic Theses and Dissertations. 4026*, 2019.

[10] S. Talukder and B. Carbunar, "When friend becomes abuser: Evidence of friend abuse in facebook," in *Proceedings of the 2017 ACM on Web Science Conference*, pp. 421–422, 2017.

[11] R. Gupta, G. Pandey, P. Chaudhary, and S. K. Pal, "Machine learning models for government to predict covid-19 outbreak," *Digit. Gov.: Res. Pract.*, vol. 1, Aug. 2020.

[12] S. K. Talukder, M. I. I. Sakib, and M. M. Rahman, "Model for e-government in bangladesh: A unique id based approach," in *2014 International Conference on Informatics, Electronics Vision (ICIEV)*, pp. 1–6, May 2014.

[13] A. A. Khan, S. Shafiq, R. Kumar, J. Kumar, and A. U. Haq, "H3dnn: 3d deep learning based detection of covid-19 virus using lungs computed tomography," in *2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 183–186, 2020.

[14] S. Talukder, M. I. I. Sakib, Z. R. Talukder, U. Das, A. Saha, and N. S. N. Bayev, "Usensewer: Ultrasonic sensor and gsm-arduino based automated sewerage management," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pp. 12–17, IEEE, 2017.

[15] S. Talukder, I. I. Sakib, F. Hossen, Z. R. Talukder, and S. Hossain, "Attacks and defenses in mobile ip: Modeling with stochastic game petri net," in *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, pp. 18–23, 2017.

[16] M. Singh and I. Mukhopadhyay, "Cyber security issues in the covid-19 times," in *Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing*, pp. 671–680, Springer, 2021.

[17] S. Talukder, "Tools and techniques for malware detection and analysis," *arXiv preprint arXiv:2002.06819*, 2020.

[18] S. k. Vangipuram and R. Appusamy, "Machine learning framework for covid-19 diagnosis," in *International Conference on Data Science, E-Learning and Information Systems 2021*, DATA'21, (New York, NY, USA), p. 18–25, Association for Computing Machinery, 2021.

[19] N. Darapaneni, A. Singh, A. Paduri, A. Ranjith, A. Kumar, D. Dixit, and S. Khan, "A machine learning approach to predicting covid-19 cases amongst suspected cases and their category of admission," in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pp. 375–380, 2020.

[20] R. A. Hallman, A. Chikkula, and T. Prioleau, "Predicting criticality in covid-19 patients," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '20, (New York, NY, USA), Association for Computing Machinery, 2020.

[21] M. R. Canales-Fiscal, R. O. López, R. Barzilay, V. Treviño, S. Cardona-Huerta, L. J. Ramírez-Treviño, A. Yala, and J. Tamez-Peña, "Covid-19 classification using thermal images: Thermal images capability for identifying covid-19 using traditional machine learning classifiers," in *Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '21, (New York, NY, USA), Association for Computing Machinery, 2021.

[22] Y. Liu and Y. Xiao, "Analysis and prediction of covid-19 in xinjiang based on machine learning," in *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*, pp. 382–385, 2020.

[23] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, "Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data," KDD '20, (New York, NY, USA), p. 3474–3484, Association for Computing Machinery, 2020.

[24] K. Feng, F. He, J. Steinmann, and I. Demirkiran, "Deep-learning based approach to identify covid-19," in *SoutheastCon 2021*, pp. 1–4, 2021.

[25] J. Vrindavanam, R. Srinath, H. H. Shankar, and G. Nagesh, "Machine learning based covid-19 cough classification models - a comparative analysis," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 420–426, 2021.

[26] A. Shamsi, H. Asgharnezhad, S. S. Jokandan, A. Khosravi, P. M. Kebria, D. Nahavandi, S. Nahavandi, and D. Srinivasan, "An uncertainty-aware transfer learning-based framework for covid-19 diagnosis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1408–1417, 2021.

[27] N. H.S. Fedal and M. Dursun, "Covid-19 detection by x-ray images and deep learning," in *2021 6th International Conference on Mathematics and Artificial Intelligence*, ICMAI 2021, (New York, NY, USA), p. 51–58, Association for Computing Machinery, 2021.

[28] M. Qjidaa, Y. Mechbal, A. Ben-fares, H. Amakdouf, M. Maaroufi, B. Alami, and H. Qjidaa, "Early detection of covid19 by deep learning transfer model for populations in isolated rural areas," in *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pp. 1–5, 2020.

[29] M. A. Rahman, M. S. Hossain, N. A. Alrajeh, and B. B. Gupta, "A multimodal, multimedia point-of-care deep learning framework for covid-19 diagnosis," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 17, Mar. 2021.

[30] A. Khan, S. Younis, and H. Algethami, "Covid-19 identification using deep neural networks," in *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, pp. 1–7, 2021.

[31] M. Zokaeinikoo, P. Kazemian, P. Mitra, and S. Kumara, "Aidcov: An interpretable artificial intelligence model for detection of covid-19 from chest radiography images," vol. 12, Oct. 2021.

[32] M. Kumar, S. Gupta, K. Kumar, and M. Sachdeva, "Spreading of covid-19 in india, italy, japan, spain, uk, us: A prediction using arima and lstm model," vol. 1, July 2020.

[33] M. Sevi and AYDIN, "Covid-19 detection using deep learning methods," in *2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, pp. 1–6, 2020.

[34] A. Channa, N. Popescu, and N. u. R. Malik, "Robust technique to detect covid-19 using chest x-ray images," in *2020 International Conference on e-Health and Bioengineering (EHB)*, pp. 1–6, 2020.

[35] Y. Li and X. He, "Covid-19 detection in chest radiograph based on yolo v5," in *2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, pp. 344–347, 2021.

[36] G. A. Shadeed, A. A. Jabber, and N. H. Alwash, "I. covid-19 detection using deep learning models," in *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*, pp. 194–198, 2021.

[37] D. Goswami, A. Law, D. Chakraborty, and A. Dey, "Convolutional neural network for prediction of covid-19 from chest x-ray images," in *CSBio '20: Proceedings of the Eleventh International Conference on Computational Systems-Biology and Bioinformatics*, CSBio2020, (New York, NY, USA), p. 104–105, Association for Computing Machinery, 2020.

[38] F. F. Kharbat, T. A. Elamsy, and N. H. Hamada, "Diagnosing covid-19 in x-ray images using hog image feature and artificial intelligence classifiers," in *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, BCB '20, (New York, NY, USA), Association for Computing Machinery, 2020.

[39] L. Huang, S. Ruan, and T. Denoeux, "Covid-19 classification with deep neural network and belief functions," in *The Fifth International Conference on Biological Information and Biomedical Engineering*, BIBE2021, (New York, NY, USA), Association for Computing Machinery, 2021.

[40] M. Xian, P. H. Carey, C. Fares, F. Ren, S.-S. Shan, Y.-T. Liao, J. F. Esquivel-Upshaw, and S. J. Pearton, "Rapid electrochemical detection for sars-cov-2 and cardiac troponin i using low-cost, disposable and modular biosensor system," in *2020 IEEE Research and Applications of Photonics in Defense Conference (RAPID)*, pp. 1–2, 2020.

[41] R. Stojanovic and A. Skraba, "Simplified open hw /sw pulse oximetry interface for purpose of covid-19 symptoms detection and monitoring," in *2021 10th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–5, 2021.

[42] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, and S. R. Steinhubl, "Wearable sensor data and self-reported symptoms for covid-19 detection," *Nature Medicine*, vol. 27, no. 1, pp. 73–77, 2021.

[43] R. Stojanović, A. Škraba, and B. Lutovac, "A headset like wearable device to track covid-19 symptoms," in *2020 9th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1–4, IEEE, 2020.

[44] A. Natarajan, H.-W. Su, and C. Heneghan, "Assessment of physiological signs associated with covid-19 measured using wearable devices," *NPJ digital medicine*, vol. 3, no. 1, pp. 1–8, 2020.

[45] T. Mishra, M. Wang, A. A. Metwally, G. K. Bogu, A. W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld, *et al.*, "Pre-symptomatic detection of covid-19 from smartwatch data," *Nature biomedical engineering*, vol. 4, no. 12, pp. 1208–1220, 2020.

[46] K.-C. Un, C.-K. Wong, Y.-M. Lau, J. C.-Y. Lee, F. C.-C. Tam, W.-H. Lai, Y.-M. Lau, H. Chen, S. Wibowo, X. Zhang, *et al.*, "Observational study on wearable biosensors and machine learning-based remote monitoring of covid-19 patients," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.

[47] T. Zhu, P. Watkinson, and D. A. Clifton, "Smartwatch data help detect covid-19," *Nature biomedical engineering*, vol. 4, no. 12, pp. 1125–1127, 2020.

[48] S. D. Bukkitgar, N. P. Shetti, and T. M. Aminabhavi, "Electrochemical investigations for covid-19 detection - a comparison with other viral detection methods," *Chemical Engineering Journal*, vol. 420, p. 127575, 2021.

[49] L. Surya and R. T. Yarlagadda, "Ai economical smart device to identify covid-19 pandemic, and alert on social distancing who measures," *International Journal of Creative Research Thoughts (IJCRT), ISSN*, pp. 2320–2882, 2020.

AUTHORS

Sajedul Talukder, Ph.D. is an Assistant Professor of Computer Science at the School of Computing at Southern Illinois University and the founding director of Security and Privacy Enhanced Machine Learning Lab (SUPREME Lab). He earned his Ph.D. in Computer Science from the Cyber Security and Privacy Research (CaSPR) Lab at Florida International University. Dr. Talukder's research interests include security and privacy with applications in online and geosocial networks, machine learning, wireless networks, distributed systems, and mobile applications. Dr. Talukder's research has been published in reputed journals and leading international conferences including Nature Scientific Reports, ACM TSC, ICWSM, ACM CHI, ACM WebSci, and has received notable media attention including from NBC 6 and Sage Research Methods. Widely cited in books and research papers, his research has been funded by leading federal agencies like National Science Foundation (NSF), Cyber Florida, Florida Center for Cybersecurity etc. He is the recipient of prestigious 2022 NSF CRII award. Dr. Talukder is serving on the editorial boards and program committees in numerous prestigious conferences and journals including IEEE COMPSAC, EAI SecureComm, ACM CCSC, and ACL ALW. He is the recipient of several research travel grants and has been invited by Facebook Research to their HQ. Prior to joining SIU, he worked as a research mentor for FIU summer research programs (such as Science without Borders, NSF-RET, NSF-REU).

Faruk Hossen is a Senior Lecturer of Computer Science at Bangabandhu Sheikh Mujibur Rahman Science of Technology University (BSMRSTU). He completed his B.Sc in Computer Science and Engineering at Bangladesh University of Engineering and Technology and is pursuing his M.Sc in Computer Science and Engineering at Khulna University of Engineering and Technology. Prior to joining BSMRSTU, he worked as a senior software engineer for 4 years at Reve Systems. Hossen's research interests include network security, Machine learning, Internet of Things (IoT), web and mobile applications, sentiment analysis, and development of embedded systems. Hossen's work has been published in renowned conferences. His current research focuses on developing mental sickness detection and recovery suggestion system for mentally sick students during the COVID-19 situation.